



Blanch, R., Horsburgh, J., Creavin, A., & Burdon, M., & Williams, C. E. M. (2019). Detection of Papilloedema Study (DOPS): rates of false positive papilloedema in the community. *Eye*, 33(7), 1073-1080.
<https://doi.org/10.1038/s41433-019-0355-9>

Peer reviewed version

Link to published version (if available):
[10.1038/s41433-019-0355-9](https://doi.org/10.1038/s41433-019-0355-9)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Nature at <https://doi.org/10.1038/s41433-019-0355-9> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Title: Detection of Papilloedema Study (DOPS): Rates of False Positive Papilloedema in the Community

Authors: Blanch RJ^{1,2,3}, Horsburgh J³, Creavin A⁴, DOPS study group^{*3,5,6}, Burdon MA³, Williams C⁴

1. Academic Department of Military Surgery and Trauma, Royal Centre for Defence Medicine, Birmingham, UK
2. Neuroscience and Ophthalmology, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK
3. Birmingham Neuro-Ophthalmology, University Hospital Birmingham NHS Trust, Birmingham, B15 2WB, UK
4. Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, UK
5. Neurometabolism, Institute of Metabolism and Systems Research, University of Birmingham, Edgbaston, B15 2TT, UK
6. University Hospitals Birmingham NHS Trust, Birmingham, UK

* DOPS study group members reviewed images and included: Mollan SP^{3,5}, Lo SY⁶, Denniston AK^{2,6}, Dulku S⁶

Corresponding Author: Maj Richard Blanch, Neuroscience and Ophthalmology, Robert Aitken Building, Vincent Drive, Birmingham, B15 2TT

blanchrj@bham.ac.uk

Tel. 07973 770061

Contributorship: RB, JH, AC, MA and WC planned the research. RB, JH, AC and CW conducted the research. RB wrote the manuscript and JH, AC, DOPS study group members, MB and CW revised the manuscript.

Synopsis: Over-diagnosis of papilloedema is common. Fundus photographic screening is highly sensitive for papilloedema detection; however, we found rates of false positive papilloedema diagnosis on fundus photography of 7-21% in hospital and community patient samples respectively.

Word Count: 2589 words

ABSTRACT

Background: Overdiagnosis of papilloedema is common and carries significant potential for morbidity from over-investigation and over-treatment. We aimed to determine the community prevalence of false positive diagnosis of papilloedema (FPE) on fundus imaging.

Methods: We evaluated fundus images from a community cross-section of 198 12-14-year-olds from the Avon Longitudinal Study of Parents and Children (ALSPAC) longitudinal cohort study database and patient images from our hospital departmental database with and without papilloedema. We asked clinicians, in isolation, to rate the subjects as a forced-choice task to “papilloedema” or “not-papilloedema” based on the fundus images alone. Raters comprised (i) 4 neuro-ophthalmologists, (ii) 4 ophthalmologists, (iii) 4 neurologists and (iv) 4 emergency medicine physicians.

Results: The prevalence of FPE in the ALSPAC population, defined as images mistaken as papilloedema by $\chi\%$ of raters (P_χ) varied from $P_{100}=0\%$ to $P_{50}=21.3\pm3.9\%$. In the hospital population, there was a lower rate of FPE, $P_{50}=7.1\pm10.8\%$. Sensitivity for papilloedema detection approached 100%, though 3 raters incorrectly labelled the same patient with unilateral disc swelling as normal, all other cases were detected by all raters.

Conclusions: Fundus photography assessment in isolation is highly sensitive but poorly specific for papilloedema detection. Using this method to screen the general population has significant potential for harm as overdiagnosis occurs, even in the hands of experienced clinicians.

INTRODUCTION

On 15th July, 2016 Honey Rose, an optometrist in Suffolk, UK, was convicted of gross-negligence manslaughter for failing to detect 8 year old Vincent Barker's papilloedema in February 2012.

Although she was subsequently cleared on appeal, the case has had a significant effect on UK optometric and ophthalmic practice [1-3]. In this case, Honey Rose conducted a routine sight-test examination of Vincent Barker, which involved direct ophthalmoscopy and fundus photography [1]. She reported that direct ophthalmoscopy was unsuccessful because of photophobia and fundus photographs demonstrating papilloedema were obtained, but Ms Rose accidentally viewed the wrong images, incorrectly concluding that the fundus appearance was normal [1]. Vincent Barker died 5 months later, in July 2012, from hydrocephalus [1].

Since the conviction of Honey Rose, neuro-imaging requests and outpatient referrals for papilloedema have increased [2, 3]. The Honey Rose case centred around the use of fundus photography for papilloedema detection. Fundus photography is routinely used in most UK optometric practices and has been advocated to improve posterior segment evaluation [4].

Because papilloedema may be the presenting feature of life-threatening conditions, its diagnosis often precipitates an urgent hospital admission, brain imaging and lumbar puncture [5]. However, it is often over-diagnosed putting patients through unnecessary treatment and invasive investigations [2, 6]. False positive diagnosis of papilloedema (FPE) may be caused by an anomalous optic disc appearance, termed pseudo-papilloedema [7], or misinterpretation of the fundus appearance.

Outpatient referrals of patients with suspected papilloedema are common in UK and USA, but few of these patients are diagnosed with papilloedema [2, 8], suggesting that the prevalence of FPE is much higher than that of papilloedema.

The prevalence of optic disc drusen (ODD), which are one cause of pseudo-papilloedema, was 0.035% in a clinical study and up to 2% in cadaveric studies [9, 10]. The prevalence of FPE on fundus

photographic screening by a single neuro-ophthalmologist of a morbidly obese population undergoing bariatric surgery was up to 2% [11]. The prevalence of FPE in the general population and comparison of accuracy between different specialists has not previously been reported.

We aimed to investigate the prevalence of FPE on fundus images in a community sample of unselected children.

METHODS

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee. The study protocol was approved by the hospital trust Research and Development Department and conformed to the tenets of the Declaration of Helsinki.

Participants providing images

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal birth cohort study of children born to mothers resident in Avon, UK who had an estimated delivery date between April 1, 1991 and December 31, 1992, including approximately 72% of eligible pregnant women [12]. The ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary at <http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>. Informed consent was obtained from all subjects and parents/guardians. Subjects included were all aged between 11.8 and 14.2 years at the time of the study visit. For the purposes of this study, the ALSPAC cohort was assumed to be representative of the general population.

Patient images were obtained from the hospital eye department fundus photographic database.

Participant information

Cross-sectional data were collected from children at age 12-14 years. Demographic information collected included: age, gender, maternal-reported ethnicity socio-economic status. Socio-economic status was categorised using the highest value for parental employment from both parents

(Standard Occupational Classification). The study visit included: autorefraction using a Canon R50 autorefractor (Canon Medical Systems, Melville, NY); best corrected LogMAR visual acuity (BCVA), height, weight and body fat percentage. Height was measured with shoes and socks removed using a Harpenden stadiometer (Holtain Ltd, Crymych, Pembrokeshire, United Kingdom) to the nearest 0.1 cm, and weight was measured by using a Tanita TBF 305 body-fat analyzer and weighing scales (Tanita UK Ltd, Yewsey, Middlesex, United Kingdom). BMI was calculated as weight (kilograms)/height (meters squared). To screen for health problems, one year after the images were taken, parents and guardians were asked to report on the health of the subject over the previous year.

For hospital patients, clinical information, including the presence or absence of papilloedema, was derived from retrospective analysis of the electronic patient records and paper charts where necessary.

Gold Standard Determination of Papilloedema vs Not Papilloedema

Our main aim was to determine the prevalence of FPE. There is no gold standard test to exclude papilloedema, although options include fundus fluorescein angiography and lumbar puncture, both of which are invasive and would be unethical to perform in a large community sample such as the ALSPAC database.

In children the main causes of papilloedema are intracranial mass lesions, which have an incidence of 2-4/100,000 and idiopathic intracranial hypertension, which is less frequent than the adult incidence of 0.9/100,000 [13, 14]. The probable number of participants with papilloedema in the ALSPAC cohort was close to zero (99% confidence interval 0.001-0.02 of 150 participants assuming a frequency of 4/100,000) and for confirmation of this, we reviewed the responses about the children's health, given in parent-completed questionnaires sent out approximately a year after the children had their fundus photographs taken. We therefore assumed that all cases in which papilloedema was called in the ALSPAC community sample were FPE.

Gold standard for the hospital patients was taken as clinical diagnosis.

Fundus Images

ALSPAC images were 45° digital retinal images centered on the macula acquired using a Topcon nonmydriatic retinal camera (Topcon TRC-NW6s, Topcon Technologies, Paramus, NJ) fitted with a Nikon D1X camera (Nikon, Tokyo, Japan). Images were available for 3350 ALSPAC participants attending a multidisciplinary data collection session. All images were reviewed by two authors (RB and CW). Bilateral images from 150 patients were selected at random from the ALSPAC database. To assess intra-observer correlation, a random selection of 10 bilateral images was duplicated.

Patient fundus images were all acquired on a Topcon TRC-50DX (Type 1A) Mydriatic Retinal Camera (with Nikon D300) after pharmacologic dilatation. Bilateral fundus images from 28 patients without papilloedema and 10 patients with papilloedema were randomly selected from the Eye Department fundus photographic database. Only images centered on the macula were included, so that they were in the same format as the ALSPAC images.

Images were excluded when the image quality was rated as inadequate to assess optic disc swelling by RB and JH.

Assessment of fundus photographs

We conducted a prospective assessment in groups of 4 senior (UK Consultant level) physicians. The groups of physicians were: neuro-ophthalmologists (NO), ophthalmologists (O), neurologists (N) and emergency medicine physicians (EM). We presented ALSPAC and patient images together in a forced choice task where we required raters to assign fundus photos to papilloedema or not papilloedema groups, based on the photographic appearance alone. Physicians were not told the source of the images and were masked to any clinical details because they were only shown fundus images in isolation and in a random order with community and hospital images mixed together.

Sample size

Sample size was calculated as previously described for observational studies as $\frac{Z^2 P(1-P)}{d^2}$ [15], where Z is normal statistic for the level of confidence (1.96 for 95%), d is the level of precision required (set to half the prevalence) and P is the expected prevalence (published FPE rates are 2-12.5% [11, 16]). To detect a 10% rate of FPE with 95% confidence, requires n=138. A 5% rate of FPE requires n=291. Preliminary discussions suggested that more than 200 images would deter clinicians from rating the images and we therefore elected to present 150 ALSPAC images.

Statistical analysis

Statistical analysis was performed in SPSS 21 (IBM Corp. Armonk, NY), except for free marginal kappa which was calculated as previously described to assess inter-rater reliability (<http://justusrandolph.net/kappa/>) [17]. Intra-rater agreement was not assessed because only 10 images were repeated and none were identified as papilloedema. Unless otherwise specified, means are displayed \pm standard error of the mean. Confidence intervals (CI) for proportions were calculated using normal approximation to the binomial distribution or the binomial distribution when the proportion mean and variance were greater than 10. . Specificity (false positives compared to true negatives) was assessed by modelling the proportion of true negative responses in the ALSPAC community sample and hospital images without papilloedema using Generalised Estimating equations with a binomial model with logit identity function and an exchangeable correlation matrix [18]. Sensitivity (true positives compared to false negatives) was assessed by modelling the proportion of true positive responses in the hospital images with papilloedema, as for specificity. The relationship between patient factors and FPE was assessed by fitting a generalised linear model (negative binomial) with continuous measurements (body mass index [BMI], body fat percentage, gestational age at birth, birthweight, age, spherical equivalent and BCVA) as covariates and gender, ethnicity and maternal social class as factors.

Proportions were compared using Pearson's χ^2 test. The rate of FPE (P) is displayed as the proportion of patients who were mistaken as having papilloedema by $\chi\%$ of raters (P_χ).

RESULTS

Prevalence of FPE

The prevalence of FPE in the ALSPAC population, defined as patients who were mistaken as having papilloedema by $\chi\%$ of raters (P_χ) varied with the value of χ (Fig. 1A). No patient was incorrectly assessed as papilloedema by all raters and only one by >90% of observers ($P_{90}=0.67$; 95% CI 0-3.7%; Fig 2A), whilst 32 patients were incorrectly assessed as PE by 50% of observers ($P_{50}=21.3$; 95% CI 14.8-27.9%), a proportion that was not significantly different between the NO, O and N groups.

In the hospital population, 2 patients were incorrectly assessed as papilloedema by >50%, giving a P_{50} of 7.1 (95% CI 0.9-23.5%; Fig. 1B).

Sensitivity and specificity papilloedema detection

Sensitivity for papilloedema detection approached 100%, although one ophthalmologist and two neurologists incorrectly labelled the same patient with unilateral disc swelling (Fig. 2B) as without papilloedema (sensitivities: NO, 100%; O, 98 \pm 2.4%; N, 95 \pm 4.7%; ED 100%).

Specificity for the assessment of papilloedema was lower, with individual specificities ranging from 42.7-100%, being lowest in the EM physicians (NO, 85 \pm 2.0%; O, 90 \pm 1.7%; N, 87 \pm 2.1%; ED, 53 \pm 3.6% $p<0.001$). Specificity was lower for the ALSPAC than the hospital images (ALSPAC 75 \pm 2.2%; hospital 87 \pm 3.1%; $p=0.007$).

Consistency of decision-making

Across all raters and images, there was 72.6% agreement, with a free-marginal kappa of 0.45 (95% CI 0.4-0.5), indicating a moderate overall consistency between raters. The free-marginal kappa was highest for neurologists (0.70 [95%CI 0.63-0.77]) and lowest for emergency medicine physicians

(0.38 [95%CI 0.30-0.46]), with ophthalmologists (0.68 [95%CI 0.61-0.75]) and neuro-ophthalmologists (0.54 [95%CI 0.46-0.62]) falling in between.

The agreement of neuro-ophthalmologists and ophthalmologists were lowered by one clinician in each group with FPE rates of 0% and 0.5% respectively.

ALSPAC population

At age 12-14, 3350 subjects had fundus images available, but these subjects did not differ in terms of visual acuity or refractive error from those who attended the study visit but had no available images (see Supplementary file, Table 1).

Compared to the 10,777 subjects in the ALSPAC cohort without available images (including clinic attenders and non-attenders), those with available photographs contained 3% more females ($p=0.003$), 1% more white patients, 6% more subjects in social class 1 and 2 (higher socioeconomic status; $p<0.001$) and 1% fewer premature children ($p<0.001$; see Supplementary file, Table 2).

Patient factors affecting FPE

There was no evidence of a relationship between FPE and patient age ($p=0.180$), gender ($p=0.582$), ethnicity ($p=0.215$), maternal social class ($p=0.824$), BMI ($p=0.993$), body fat percentage ($p=0.624$), gestational age at birth ($p=0.548$), mother's age at birth (0.707), birth weight ($p=0.545$), refraction ($p=0.212$) or BCVA ($p=0.651$; Fig. 3). All except one ALSPAC subject were reported as being healthy in the subsequent year after the images were taken and that one subject was judged to have papilloedema by one ED physician only.

DISCUSSION

The P_{50} rate of FPE in our community-based sample was 21.3%, suggesting a very high potential for asymptomatic members of the general population to be referred for papilloedema investigations based on fundus photography screening alone.

Our sample size calculations suggested that our study had 95% power to detect an FPE rate greater than 10%, less than the 21.3% detected, suggesting that we were adequately powered to define this prevalence with a precision less than 5%.

The forced choice task in the absence of clinical information limits the generalisability of the results to experienced neuro-ophthalmologists, who are very likely to use additional information to aid their decision-making. However, the finding of apparent papilloedema on fundus examination should usually precipitate further investigation, whether additional features are present or not, and less experienced and non-medical professionals may be less able to use additional clinical features from history and examination in their decision-making: thus in the community or Emergency Department setting practitioners often face a forced choice of whether or not to refer.

Subtle papilloedema is difficult to distinguish from pseudo-papilloedema, and our data do not suggest a solution beyond the utility of a second opinion. Optical coherence tomography (OCT) currently has limited utility to differentiate, as increased retinal nerve fibre layer thickness has been reported in both papilloedema and pseudo-papilloedema [19, 20]. Fluorescein angiography is most sensitive, but is invasive and often not readily available in the community [21]. The clinician may also assess spontaneous venous pulsation (SVP) at the optic disc, which strongly suggests that papilloedema is absent but is unfortunately less common in eyes with anomalous optic discs [22].

The rate of FPE may be lower when patients are assessed on a slit lamp or with a direct ophthalmoscope, but this was not the case in the FOTO-ED studies, where clinical examination alone did not usually help distinguish FPE after fundus photographic screening by neuro-ophthalmologists [personal communication – Dr. Beau Bruce, Emory University, Atlanta, GA USA, 2018].

In our study, the detection of papilloedema by fundus photography was extremely sensitive, even among EM physicians, consistent with previous reports [16, 23-25]. That three raters missed the same case of unilateral papilloedema suggests a need to highlight that papilloedema may be asymmetric [26]. The exponential reduction in FPE rates with a requirement for increasing

agreement, suggests that second and third opinions from ophthalmology or neurology colleagues may help differentiate papilloedema from FPE.

In determining the sensitivity of papilloedema detection, taking the gold standard as clinical diagnosis in the hospital patients and questionnaire assessment plus assumed normality in the community sample limits generalisability, because one or two missed cases of papilloedema in the community or hospital samples would greatly reduce observed sensitivity. However, we did not design the study to assess the sensitivity of fundus photographic screening for ophthalmic pathology, which has been previously assessed in the FOTO-ED studies [16, 23-25], which found a FPE rate of 2-12.5% after EM physician review of fundus photography in patients for whom fundus examination was indicated and a sensitivity for papilloedema detection of 83% or less [16]. These studies did not systematically assess the rate of FPE in neuro-ophthalmologists' image assessments, although this was very low [personal communication – Dr. Beau Bruce, Emory University, Atlanta, GA USA, 2018]. This difference from our study may reflect different context and approaches (FOTO-ED involved real clinical decisions) or that the FOTO-ED neuro-ophthalmologists would all have fallen towards the low-FPE end of our panel of raters.

The agreement among individual physicians on which patients had FPE was moderate to substantial (kappa 0.45 -0.7), greater than a previous study, which found a kappa range 0.17–0.43 for non-fluorescein imaging modalities [21], suggesting common decision-making strategies. None of our 10 cases of papilloedema were mild (Frisén 1) in both eyes and we therefore cannot comment on the level of agreement between observers on subtle true disc swelling. However, all FPE cases appeared to be no more than subtle papilloedema, so the disagreement between observers was on what constituted subtle disc swelling.

The hospital population had a significantly lower FPE rate (and higher specificity of a papilloedema diagnosis) than our community-based sample, though at 7.1%, this still leaves a very high potential for the hospital population to be referred for papilloedema investigations. Possible explanations for

this difference include that the community-based cohort were aged 12-14 years, whilst the hospital database included predominantly adults. However, there was no significant effect of age on FPE assessment in our data and in a recent study of methods to diagnose papilloedema, FPE was more common in patients under 12 years [21], suggesting that our cohort of patients over 12 years should not have an age-related increase in the FPE rate.

Our assumption that the ALSPAC subjects did not have papilloedema was based on the very low population rate (<0.01%) and the lack of reported health problems by the subjects' parents and guardians one year later [13, 14]. There were no cases of frank papilloedema in the included cases or in any of the other 3350 images viewed by RB and CW whilst preparing the study and there was no relationship between FPE and other known associations with papilloedema such as BMI and body fat percentage, suggesting that this assumption was sound. Were this assumption violated, two cases of true papilloedema in the community sample would change the observed rate of FPE or specificity by only 1.3%.

The P_{50} rate of FPE was $21.3 \pm 3.9\%$ for the ALSPAC images and $7.1 \pm 10.8\%$ for the hospital images, suggesting that screening of the general population by fundus imaging has significant potential for harm in terms of over-diagnosis of papilloedema with pressure on secondary care services and morbidity from investigations and great but unnecessary anxiety caused to patients and families. However, the high sensitivity for papilloedema detection in all groups including EM physicians supports its targeted use in patients for whom fundus examination is indicated to exclude papilloedema.

FUNDING

The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors who will serve as

guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website.

This study was supported in part by NIHR Doctoral (AC; DRF-2015-08-018) and Senior (CW; SRF-2015-08-005) Research Fellowships. Dr Creavin also works as part of the Medical Research Council Integrative Epidemiology Unit at the University of Bristol, which is funded by the MRC grant code: MC_UU_12013/3.

ACKNOWLEDGEMENTS

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

DECLARATION OF COMPETING INTERESTS

No authors contributing have a competing or conflicting interest in the subject matter.

REFERENCES

1. White J. Criminal Trial of Optom. *Optometry Today*, 2016,
<https://www.aop.org.uk/ot/professional-support/clinical-and-regulatory/criminal-trial-of-optometrist-begins>.
2. Dahlmann-Noor AH, Adams GW, Daniel MC, Davis A, Hancox J, Hingorani M, *et al*. Detecting optic nerve head swelling on ultrasound and optical coherence tomography in children and young people: an observational study. *Br J Ophthalmol* 2018; 102: 318-322.
3. Poostchi A, Awad M, Wilde C, Dineen RA, Gruener AM. Spike in neuroimaging requests following the conviction of the optometrist Honey Rose. *Eye* 2017; 32: 489-490.
4. Dabasia PL, Edgar DF, Garway-Heath DF, Lawrenson JG., A survey of current and anticipated use of standard and specialist equipment by UK optometrists. *Ophthalmic Physiol Opt* 2014; **34**: 592-613.
5. Mollan SP, Markey KA, Benzimra JD, Jacks A, Matthews TD, Burdon MA., *et al*. A practical approach to, diagnosis, assessment and management of idiopathic intracranial hypertension. *Pract Neurol* 2014; **14**: 380-390.
6. Fisayo A, Bruce BB, Newman NJ, Biousse V., Overdiagnosis of idiopathic intracranial hypertension. *Neurology* 2016; **86**: 341-350.
7. Rosenberg MA, Savino, P.J., Glaser, J.S. A clinical analysis of pseudopapilledema. I. Population, laterality, acuity, refractive error, ophthalmoscopic characteristics, and coincident disease. *Arch Ophthalmol* 1979; **97**: 65-70.
8. Kovarik JJ, Doshi PN, Collinge JE, Plager DA. Outcome of pediatric patients referred for papilledema. *J AAPOS* 2015; **19**: 344-348.
9. Friedman AH, Gartner, S. Modi, S.S. Drusen of the optic disc. A retrospective study in cadaver eyes. *Br J Ophthalmol* 1975; **59**: 413-421.
10. Wilkins JM, Pomeranz H.D. Visual manifestations of visible and buried optic disc drusen. *J Neuroophthalmol*, 2004; **24**: 125-129.

11. Krispel CM, Keltner JL, Smith W, Chu DG, Ali MR. Undiagnosed papilledema in a morbidly obese patient population: a prospective study. *J Neuroophthalmol* 2011; **31**: 310-315.
12. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al., Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013; **42**: 111-127.
13. Kaatsch P, Rickert CH, Kuhl J, Schuz J, Michaelis J. Population-based epidemiologic data on brain tumors in German children. *Cancer*, 2001. **92**(12): p. 3155-64.
14. Rigi M, Almarzouqi SJ, Morgan ML, Lee AG. Papilledema: epidemiology, etiology, and clinical management. *Eye Brain*, 2015; **7**: 47-57.
15. Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. *Gastroenterol Hepatol Bed Bench* 2013; **6**: 14-17.
16. Bruce BB, Thulasi P, Fraser CL, Keadey MT, Ward A, Heilpern KL, et al. Diagnostic accuracy and use of nonmydriatic ocular fundus photography by emergency physicians: phase II of the FOTO-ED study. *Ann Emerg Med* 2013; **62**: 28-33 e1.
17. Randolph JJ. Free-Marginal Multirater Kappa (multirater K[free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. *Paper presented at the Joensuu Learning and Instruction Symposium*; Oct 14-15, 2005; Joensuu, Finland.
<https://files.eric.ed.gov/fulltext/ED490661.pdf>.
18. Genders TS, Spronk S, Stijnen T, Steyerberg EW, Lesaffre E, Hunink MG. Methods for calculating sensitivity and specificity of clustered data: a tutorial. *Radiology* 2012; **265**:910-916.
19. Bassi ST, Mohana KP. Optical coherence tomography in papilledema and pseudopapilledema with and without optic nerve head drusen. *Indian J Ophthalmol* 2014; **62**: 1146-1151.
20. Fard MA, Fakhree S, Abdi P, Hassanpoor N, Subramanian PS. Quantification of peripapillary total retinal volume in pseudopapilledema and mild papilledema using spectral-domain optical coherence tomography. *Am J Ophthalmol* 2014; **158**: 136-143.

21. Chang MY, Velez FG, Demer JL, Bonelli L, Quiros PA, Arnold AC, et al. Accuracy of Diagnostic Imaging Modalities for Classifying Pediatric Eyes as Papilledema Versus Pseudopapilledema. *Ophthalmology* 2017; **124**: 1839-1848.
22. Hedges TR Jr, Baron EM, Hedges TR, 3rd, Sinclair SH. The retinal venous pulse. Its relation to optic disc characteristics and choroidal pulse. *Ophthalmology* 1994; **101**: 542-547.
23. Bruce BB, Lamirel C, Biousse V, Ward A, Heilpern KL, Newman NJ, et al. Feasibility of nonmydriatic ocular fundus photography in the emergency department: Phase I of the FOTO-ED study. *Acad Emerg Med* 2011; **18**: 928-33.
24. Lamirel C, Bruce BB, Wright DW, Delaney KP, Newman NJ, Biousse V. Quality of nonmydriatic digital fundus photography obtained by nurse practitioners in the emergency department: the FOTO-ED study. *Ophthalmology* 2012; **119**: 617-624.
25. Lamirel C, Bruce BB, Wright DW, Newman NJ, Biousse V. Nonmydriatic digital ocular fundus photography on the iPhone 3G: the FOTO-ED study. *Arch Ophthalmol* 2012; **130**: 939-940.
26. Bidot S, Bruce BB, Saindane AM, Newman NJ, Biousse V. Asymmetric papilledema in idiopathic intracranial hypertension. *J Neuroophthalmol* 2015; **35**: 31-36.

Figure legends

Figure 1A. FPE rate in the ALSPAC population at different levels of inter-observer agreement (P_{0-100}) for all observers and the different specialty groups. Error bars are 95% confidence intervals. **B.** FPE rate in the hospital population at different levels of inter-observer agreement (P_{1-16}) for all observers and the different specialty groups. Error bars are 95% confidence intervals.

Figure 2. Color fundus images. **A.** ALSPAC photo classified as papilloedema (FPE) by 15/16 clinicians. **B.** Case of asymmetric papilloedema missed by 3 clinicians.

Figure 3. Scatter charts showing the relationship between FPE and patient factors. Bubble size relates to the number of duplicate points (equal x and y values). None of the variables showed any evidence of a relationship with FPE. **A.** Body mass index. **B.** Gestational age at birth. **C.** Body fat percentage. **D.** Spherical equivalent refraction. **E.** Birthweight. **F.** Age at assessment.